Chapter 4: Analysis of Data from the Arnot 001, 003 and 004 Discharges

The Arnot mine site is located in Tioga County, Pennsylvania in the northeastern portion of the bituminous coal region. The Arnot discharges are from an abandoned underground mine on the Bloss (B) coal seam, which is the subject of a hydrogeologic study by Duffield (1985). The relationships between flow and water quality parameters of the Arnot mine site are also described in Smith (1988) and Hornberger et al. (1990). A map of the Arnot site is shown in Figure 4.1. The data set for the Arnot site contains 82 samples from each of the 3 mine drainage discharges for the time period from January 28, 1980 to August 14, 1983.

It is advisable to examine the distribution of the missing values because they will lead to difficulties as the analytical (statistical) tools get more sophisticated. In particular, time series analysis demands observations at regular time intervals. On the other hand, it is impractical to expect that there will be no missing values, because during a storm event, the sampling location may become inaccessible for various time intervals. It is best, therefore, to recommend time interval limits for the period in which a sample may be taken, and which, for statistical analysis, will be considered to be within the time interval (e.g., any time within a two week period will be assigned as an observation taken 2 weeks apart at the mid-point of the time interval). In any case it is advisable to examine the data carefully before attempting a quantitative analysis. Therefore, it is recommended that a graph of discharge (and/or other variables) against time be prepared and examined carefully to determine the distribution of missing values, position of the extremes, etc.

A typical example is illustrated in Figure 4.2 which is a plot of log (base 10) of flow versus time for all three point sources from the same mine. The flow for Arnot 001 is usually the largest followed by Arnot 004, then Arnot 003. All three show the same general pattern of variation.

The samples were supposed to be taken at 14 day intervals but, in practice, the intervals vary from 1 day up to 40 days. All intervals equal to or exceeding 20 days are accented in Figure 4.2. These longer intervals include, of course, many missing values. When a time series model is fitted to these data, they are “forced” into equal interval status. The effect of these departures from equal intervals is to suppress any seasonal periodicity that may be present.

It may be observed that in 1980 the runoff occurred in March, April and May; in 1981 in March and April; in 1982 in March and June; and in 1983 in April and May. These variations tend to suppress any seasonal effect in the occurrence of extreme values.

Figure 4.1: Map of Arnot Mine Site

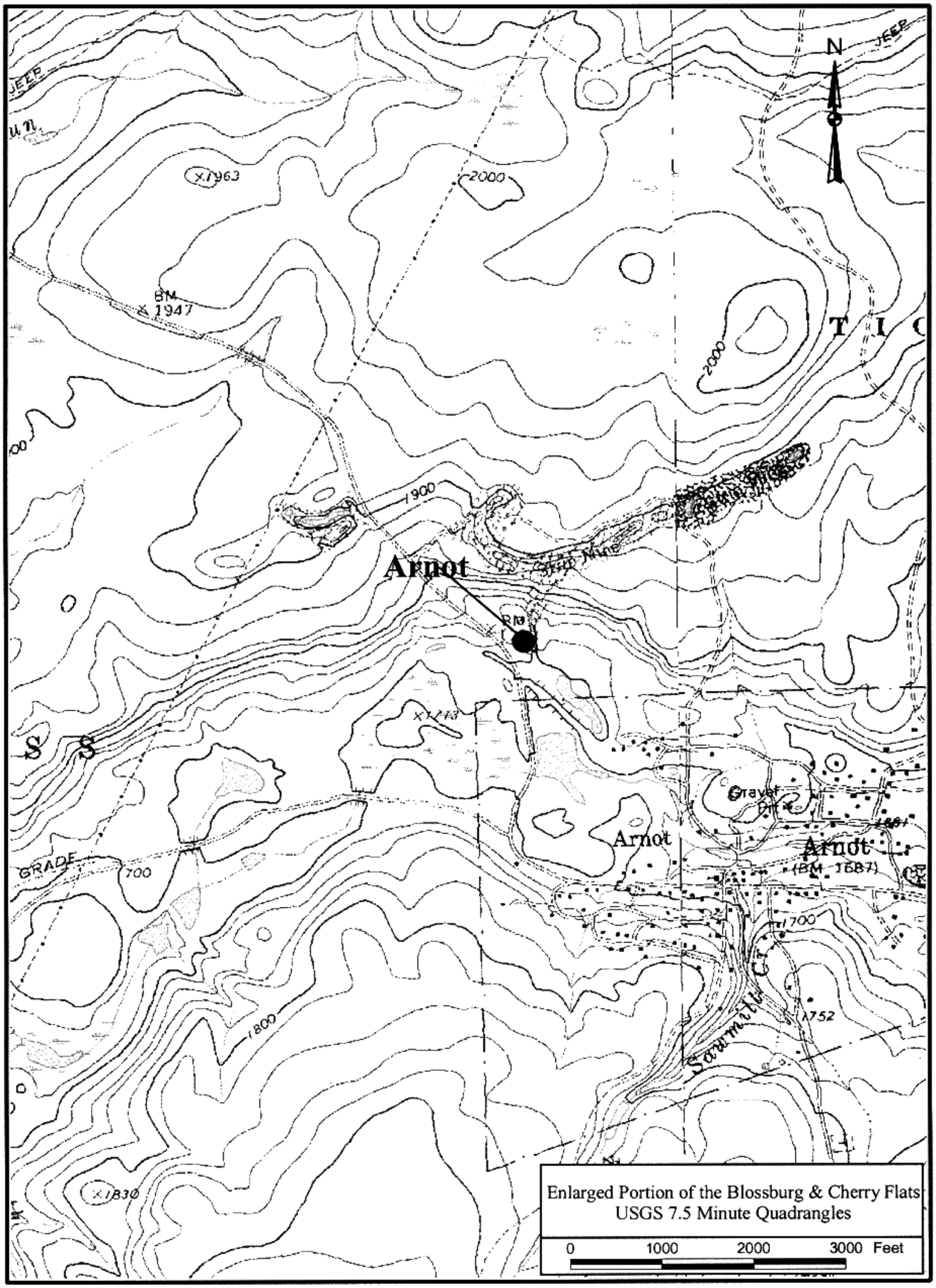
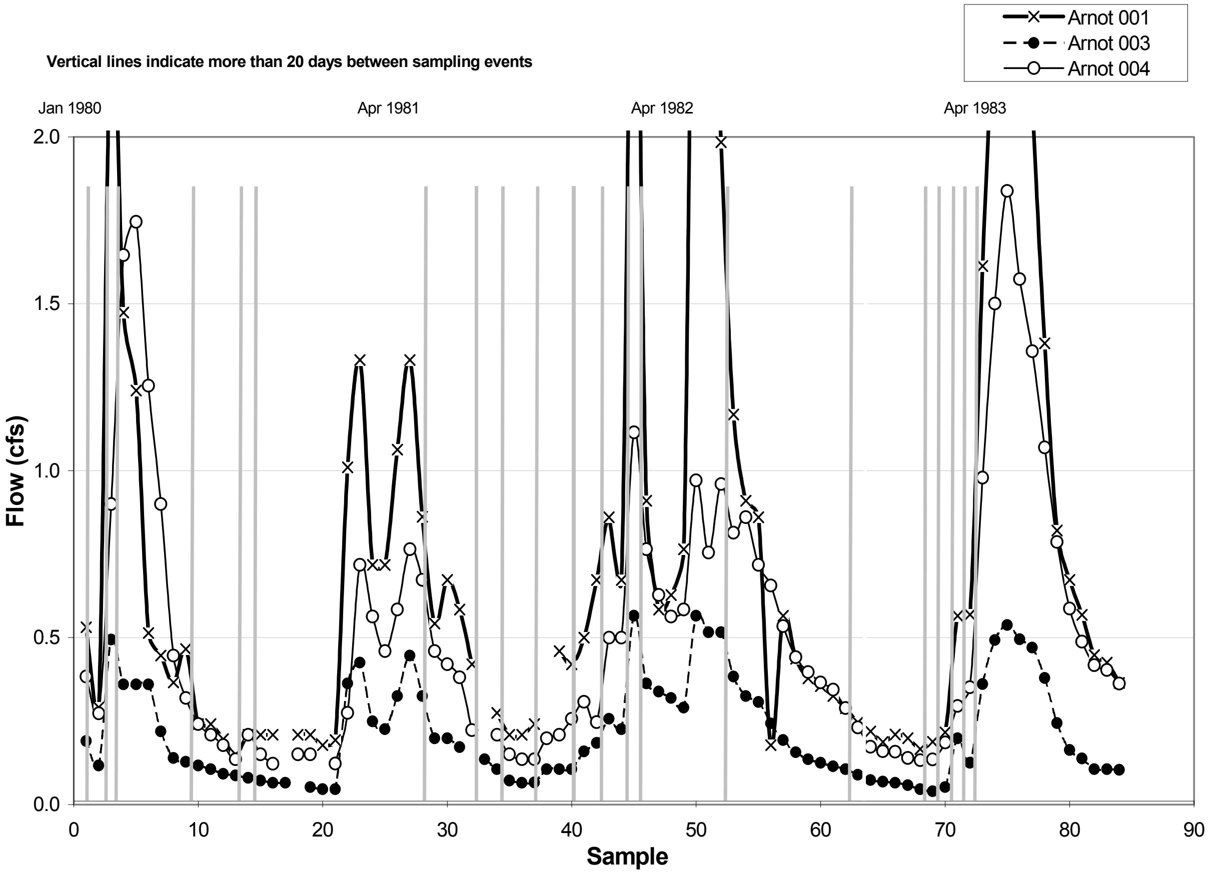


Figure 4.2: Log Flow vs. Time (Arnot 001, 003, and 004) Procedures to Adjust the Data Set for Missing Data



In some cases, it will be advantageous to insert some suitable value in place of the missing observation and the procedures for selection of a suitable value can differ. One such approach is to insert the mean value for the series or, if the frequency distribution is somewhat skewed (asymmetric), the median may be more representative.

There are also smoothing procedures varying from simple ones, such as the average of a pair of values on either side of the missing observation, through running averages using any of several larger sets of numbers. These, smoothing procedures, are described in Velleman and Hoaglin (1981, Chapter 6) and (Cleveland, 1979).

A typical, but rather elaborate example, specifically designed for time series analysis, is described by Damsleth (1986). This example begins with “simple linear interpolation between observations preceding and following the gap” then identifying and estimating a univariate time series model for the “adjusted series” which, in turn, yields “optimal estimators using the model.” The new series is used to build a transfer function model between two series (such as acidity and flow) and calculating new optimal values which are in turn used to estimate new model parameters (Damsleth, p. 46-47). The conclusions reached by Damsleth (p. 47) are: “The various steps in the process gave only small changes in the estimates for missing values, and the model and parameter estimates were almost unaffected….”.

It should be clear that missing observations can be a very difficult problem. Another aspect of this “data massaging” procedure arises when attempts are made to reduce the magnitude of the error of residuals when fitting a time series model. In a series of flow observations, for example, there maybe some extremely large values that arise from unusual events (e.g., heavy rainfall, perhaps persisting for several days, sudden water run-off from snow melt, etc.). These “natural” events of limited duration can increase the residual error quite seriously and usually do not represent persistent increased contamination. In the series of mine drainage data examined in this chapter, these unusually large values are often associated with missing data. This means that if one inserts a very small value (near zero) for the missing value, the entire range in parameter values occurs within a short period. It is advisable to reduce this wide range, first by not using low values for zero or missing values but by using one of the procedures described above. Secondly, the extreme high values should be smoothed out (i.e., large variance, and wide confidence limits which tend to be insensitive to large departures in the data). The effects of these adjustments may be estimated by running the series, after removing the zero values, both with the original extreme values and with the extremes adjusted by some form of smoothing.

In comparing the results of the more sophisticated smoothing technique described by Damsleth with other “quick and dirty” techniques, it was found that the changes were not very different. Therefore, it was concluded that elaborate smoothing procedures are unnecessary for mine drainage data sets.

Univariate Analysis

The analysis commences with the summary statistics displayed in Tables 4.1a to 4.1c. The number of samples (N) for each variable is listed first, followed by the number of missing values (N*). The statistical summary then follows with values for the arithmetic mean, median, trimmed (10%) mean, standard deviation, standard error of the mean, minimum, maximum, and the first and third quartiles. A convenient procedure for comparing variabilities among different variables, and among the same variables from different sources, is by means of the Coefficient of Variation (CV) where CV% = (standard deviation / mean) *100 expressed in percent. The values are displayed in Table 4.2 for convenient comparisons. For all three Arnot sources, pH has the smallest variability (around 4%), whereas, discharge has the largest variability (Arnot 1: CV=112%, Arnot 3: CV=70.0%, Arnot 4: CV=78.1%).

Table 4.1a: Summary Statistics for Arnot 001 Data

	N	N*	Mean	Median	Trimmed Mean	Standard Deviation	Standard Error of the Mean
pH	81	0	4.8505	4.8400	4.8479	0.2221	0.0247
Temperature	67	14	9.448	9.100	9.403	1.424	0.174
Discharge	81	0	0.7961	0.5000	0.6747	0.8955	0.0995
Acidity	81	0	20.04	16.00	19.42	11.26	1.25
Alkalinity	81	0	6.457	5.000	5.918	5.480	0.609
Total Iron	81	0	0.21111	0.20000	0.21096	0.07583	0.00843
Ferrous Iron	81	0	0.11728	0.10000	0.11507	0.07872	0.00875
SO_4	81	0	173.23	177.00	173.22	44.05	4.89
Ca	75	6	109.52	111.00	109.73	22.76	2.63
Mg	75	6	86.03	82.00	85.76	24.87	2.87
Mn	75	6	1.7104	1.6200	1.6776	0.6666	0.0770
Al	72	9	1.425	1.045	1.384	0.982	0.116

	Minimum	Maximum	First Quartile	Third Quartile	Coefficient of Variation
pH	4.2000	5.4500	4.6800	5.0200	4.5
Temperature	7.000	12.900	8.400	10.000	15.1
Discharge	0.0100	5.0910	0.2300	0.8615	112.0
Acidity	3.00	64.00	11.00	28.00	56.2
Alkalinity	0.000	37.000	3.000	8.000	84.8
Total Iron	0.00000	0.40000	0.20000	0.25000	35.9
Ferrous Iron	0.00000	0.30000	0.10000	0.20000	67.1
SO_4	66.00	277.00	140.50	201.50	25.4
Ca	66.00	152.000	93.00	127.00	20.8
Mg	31.00	145.000	69.00	104.00	28.9
Mn	0.5400	3.9500	1.2800	2.0300	39.0
Al	0.100	3.640	0.602	2.277	68.9

Table 4.1b: Summary Statistics for Arnot 003 Data

	N	N*	Mean	Median	Trimmed Mean	Standard Deviation	Standard Error of the Mean
pH	82	0	3.2782	3.265	3.2727	0.1095	0.0121
Temperature	67	15	8.551	8.600	8.548	0.916	0.112
Discharge	82	0	0.2157	0.1610	0.2671	0.1509	0.0167
Acidity	82	0	86.37	84.50	85.7	22.55	2.49
Total Iron	82	0	1.0963	1.1000	1.0919	0.2843	0.0314
Ferrous Iron	82	0	0.3610	0.3000	0.3405	0.2340	0.0258
SO_4	82	0	168.99	165.00	168.66	43.79	4.84
Ca	75	7	59.75	61.00	59.52	11.69	1.35
Mg	75	7	73.60	70.00	72.49	23.00	2.66
Mn	77	5	3.203	2.760	3.110	1.338	0.152
Al	73	9	5.079	4.680	5.060	2.213	0.259

	Minimum	Maximum	First Quartile	Third Quartile	Coefficient of Variation
pH	3.0400	3.7000	3.2100	3.3325	3.3
Temperature	6.200	11.700	8.100	9.000	10.7
Discharge	0.04	0.5650	0.1010	0.3282	70.0
Acidity	42.00	151.00	67.75	104.00	26.1
Total Iron	0.3000	2.0000	0.9000	1.2000	25.9
Ferrous Iron	0.0000	1.5000	0.2000	0.4000	64.8
SO_4	85.00	262.00	134.00	211.25	25.9
Ca	38.00	90.00	49.00	69.00	19.5
Mg	38.00	142.00	55.00	89.00	31.2
Mn	1.540	6.900	2.040	4.350	41.7
Al	0.700	9.440	3.400	6.960	43.6

Table 4.1c: Summary Statistics of Arnot 004 Data

	N	N*	Mean	Median	Trimmed Mean	Standard Deviation	Standard Error of the Mean
pH	81	0	3.2794	3.2800	3.2675	0.1409	0.0157
Temperature	67	14	8.466	8.600	8.487	0.906	0.111
Discharge	81	0	0.5307	0.4030	0.4887	0.4143	0.0460
Acidity	81	0	96.99	96.00	95.85	26.61	2.96
Total Iron	81	0	1.2630	1.200	1.243	.418	0.0464
Ferrous Iron	81	0	0.4198	0.3000	0.3973	0.2638	0.0293
SO_4	80	1	171.80	166.50	170.79	39.04	4.36
Ca	75	6	54.293	54.000	54.164	8.022	0.926
Mg	75	6	67.68	65.00	67.27	18.52	2.14
Mn	75	5	2.714	2.445	2.637	0.979	0.112
Al	73	8	6.453	5.900	6.317	2.590	0.303
Log Discharge	81	0	-0.3954	-0.3947	-0.4024	0.3266	0.0363
Ferric Iron	81	0	0.843	0.800	0.845	0.382	0.043

	Minimum	Maximum	First Quartile	Third Quartile	Coefficient of Variation
pH	3.0000	3.9400	3.1900	3.3350	4.3
Temperature	6.100	10.700	8.100	9.000	10.7
Discharge	0.1220	1.8380	0.2090	0.7365	78.1
Acidity	62.00	168.00	73.00	121.00	27.4
Total Iron	0.600	2.8	0.900	1.500	33.1
Ferrous Iron	0.0000	1.4	0.2500	0.5000	62.8
SO_4	86.00	268.00	143.00	200.00	22.7
Ca	39.000	79.000	49.000	60.000	14.8
Mg	17.00	110.00	54.00	75.00	27.4
Mn	1.200	6.500	1.987	3.247	36.1
Al	0.710	13.560	4.325	8.350	40.1
Log Discharge	-0.9136	0.2643	-0.6799	-0.1330	82.6
Ferric Iron	0.000	1.700	0.600	1.100	5.1

Table 4.2: Coefficient of Variation (%)

Variable	Arnot 001	Arnot 003	Arnot 004
pH	4.5	3.3	4.3
Temperature	15.1	10.7	10.7
Flow	112.0	70.0	78.1
Log (Discharge)	-	-	82.6
Acid	56.2	26.1	27.4
Alkalinity	84.8	-	-
Total Iron	35.9	25.9	33.1
Ferrous Iron	67.1	64.8	62.8
Ferric Iron	-	-	5.1
SO_4	25.4	25.9	22.7
Ca	20.8	19.5	14.8
Mg	28.9	31.2	27.4
Mn	39.0	41.7	36.1
Al	68.9	43.6	40.1

The same CV order of magnitude is maintained by each variable in each of the three sources. Log discharge does nothing to reduce the relative variation (CV) as can be seen from the value for Arnot 004 (82.6%). Discharge is highest in Arnot 001, moderate in 004, and lowest in 003. The coefficient of variation reflects this order and suggests that this parameter varies in proportion to its absolute value (heteroscedastic), again reinforcing that the appropriate transformation is to logarithms.

The majority of the variables in the histogram-like displays of data from Arnot 001 are symmetrical, such as sulfate shown in Figure 4.3. The most asymmetric is discharge which is seen in Figure 4.4. When this variable is transformed to logarithms it becomes symmetrical.

Figure 4.3: Stem-and-leaf of Sulfate (Arnot 001)

```
N = 81

Leaf Unit = 10

1         0         6

5         0         9999

9         1         0011

17        1         22223333

30        1         4444444555555

(13)      1         6666677777777

38        1         88888888888889999

21        2         000001111

12        2         2233

8         2         4445555

1         2         7
```

Figure 4.4: Stem-and-leaf of Discharge (Arnot 001)

```
N = 81

Leaf Unit = 0.10

40        0         0111100000222222222222222223333344444444

(24)      0         555555555566666777888899

17        1         0123334

10        1         69

8         2         13

6         2         69

4         3         014

1         3

1         4

1         4

1         5         0
```

The Arnot 003 and 004 data are substantially similar to that of Arnot 001. The histogram of pH data for the Arnot 003 discharge is very symmetrical, as shown in Figure 4.5, as is the histogram of sulfate data for the Arnot 004 discharge shown in Figure 4.6. Flow measurement data of the Arnot 004 discharge are asymmetric and positively skewed, as shown in Figure 4.7.

Figure 4.5: Stem-and-leaf of pH (Arnot 003)

```
N = 82

Leaf Unit = 0.010

1          30         4

2          30         7

5          31         134

17         31         555667888999

36         32         0011111112234444444

(15)       32         555666777789999

31         33         001111222234

19         33         556666778

10         34         1122

6          34         679

3          35         0

2          35         7

1          36

1          36

1          37         0
```

Figure 4.6: Stem-and-leaf of Sulfate (Arnot 004)

```
N = 80

Leaf Unit = 1.0          N* = 1

1          0          8

3          11         00

16         1          2222333333333

33         1          44444444444555555

(18)       1          666666666666667777

29         1          88888999

21         2          000000111

12         2          2222233

5          2          455

2          2          66
```

Figure 4.7: Stem-and-leaf of Acidity (Arnot 004)

Figure 4.7 : Stem-and-leaf of Acid.

```
N = 81

Leaf Unit = 1.0

13      6       2444456677799

32      7       0011223333445677899

38      8       001238

(6)     9       256778

37      10      0000337

30      11      112245558

21      12      112445678

12      13      01123677

4       14      05

2       15      2

1       16      8
```

Bivariate Analysis

The relationships between log discharge and every other parameter are similar (i.e., inverse and approximately linear). That is, as discharge increases in volume the amount of each variable, calcium, magnesium, manganese and aluminum, decreases, or in high flows the concentration is diluted. A good example of this relationship is the plot of manganese versus flow for the Arnot 001 discharge, shown in Figure 4.8.

Figure 4.8: Plot of Manganese vs. Log Flow (Arnot 001)

```
MTB > PLOT C6 VS C8
```

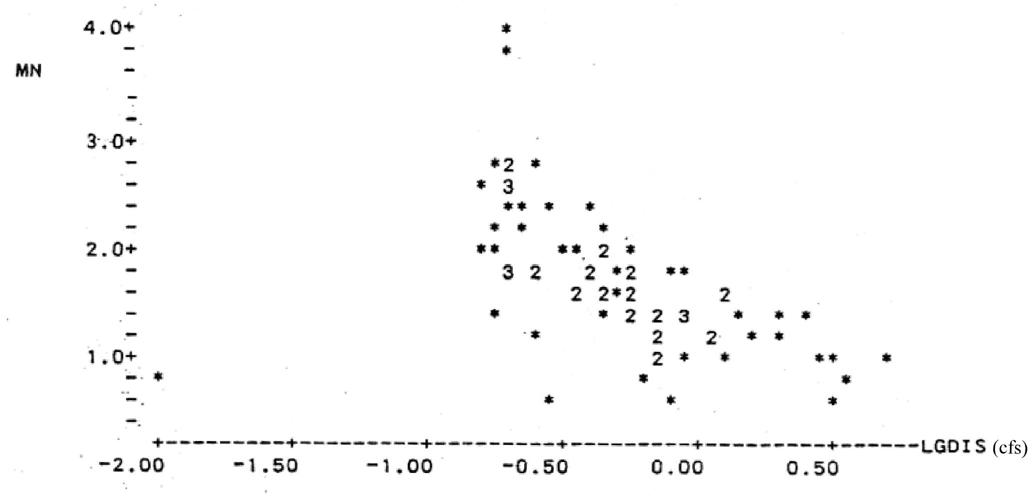
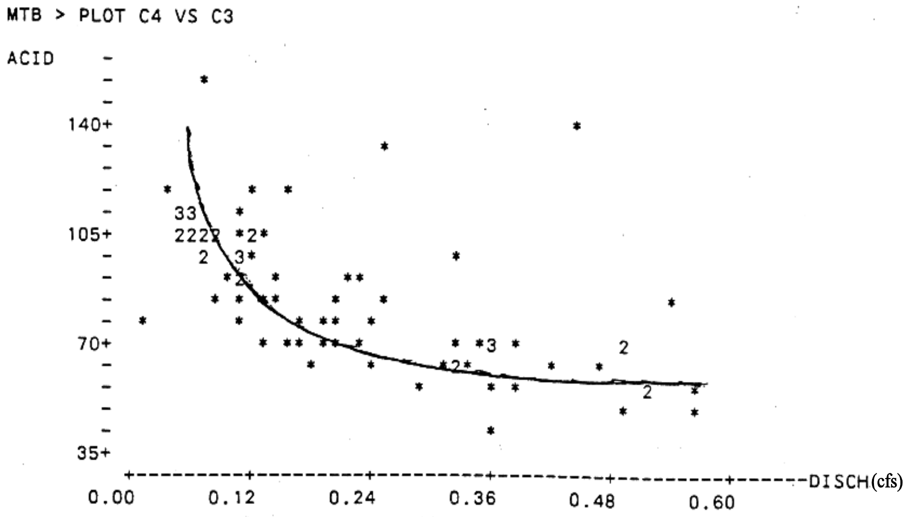
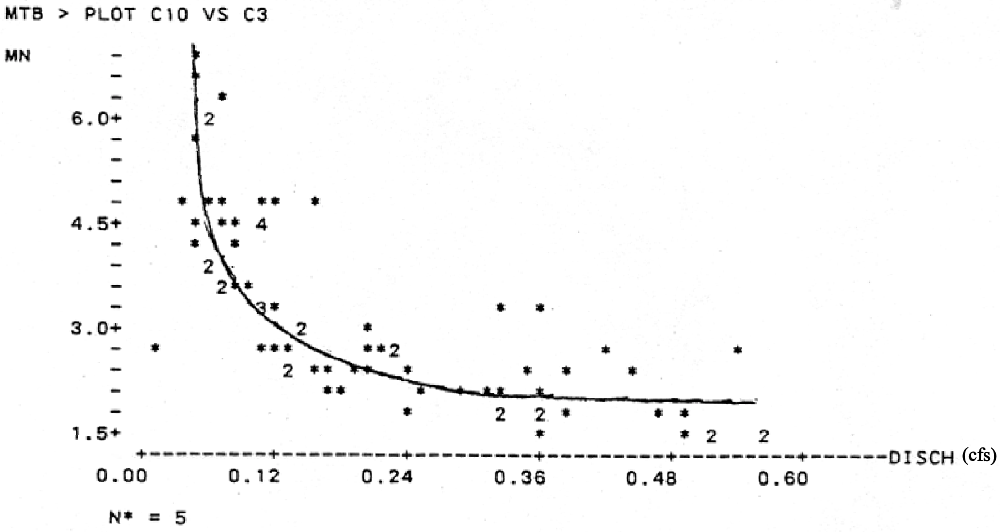


Figure 4.9: Plot of Acidity vs. Flow (Arnot 003)



MTB > PLOT C4 VS C3

Figure 4.10: Plot of Manganese vs. Flow (Arnot 003)



MTB > PLOT C10 VS C3

N* = 5

For Arnot 003, acidity vs. discharge possesses a clear, curvilinear association (Figure 4.9), which would become inversely linear if discharge was expressed in logs. Cross-correlation of these variables had maximum association of –0.648 at zero lag, or about 42% of the variation (r^2)is common to both variables. Sulfate, manganese and aluminum vs. discharge also showed this same curvilinear association of dilution with increasing flow. The example of manganese is seen in Figure 4.10.

A plot of sulfate versus acidity from the Arnot 004 discharge data showed the expected positive association but again the scatter around a straight line is very large. The expected association of calcium and magnesium is extremely weak. Any relationship between manganese and total iron is obscured by an extreme value in iron. It seems somewhat strange that the data from Arnot 004, which is located between Arnot 001 and 003, should present such a confused picture of these bivariate relationships relative to those of Arnot 001 and 003 data; possibly Arnot 004 contains more outliers than 001 or 003.

Time Series Analysis

A qualitative time series analysis was performed by plotting successive variables against (equal interval) time periods. It is convenient to start with the variable discharge (flow) for Arnot 003 (Figure 4.11a) which may be compared with the same plot on a much larger scale (Figure 4.2). The four maxima (peaks) are quite striking in both graphs. Since the date of the first observation is January 28, 1980, the first peak is in March (1980), marked in the graph by the number 3; the numbers in Figure 4.11a go from 1 to 10 (=0) and then start at 1 again and so on for each cycle of 10. The next peak is 22 (March, 1981) followed closely by another at 26 (May, 1981). Subsequent peaks occur at 43 (March, 1982), 48 (June, 1982) and then 73 (April, May, 1983). Suppose there existed an annual cycle (i.e., 26 observations, one every two weeks) then, starting with March = 3, the next peak should be 29, then 55, 81, etc. Missing observations (see Figure 4.2) and peak discharges at varying intervals, not equal annual cycles, make a seasonal pattern obscure.

Using discharge as the base which controls the concentration of acidity for example, one would expect pH to also show similar cycles in Figure 4.11b. Instead, the first peak and the following double peak are similar to those shown by discharge, but the peak at 40 is not. There is a peak at 48 in both plots but then the pH declines and stays below its mean throughout the subsequent series; there is no sign of the discharge peak at 73. The scatter diagram of pH vs. discharge showed no relationship.

The relationship between acidity and time (Figure 4.11c), tends to be inversely related to the relationship between discharge and time, i.e., the peaks of discharge coincide with the minima (maximum dilution) of acidity. This is supported by the scatter diagram between acidity and discharge (Figure 4.9). There is a slight tendency for this to be true of total iron (Figure 4.11d) but there was no sign of such a relationship in the scatter diagram of iron vs. discharge.

Sulfate, as expected from its scatter plot against discharge shows inverse relationships in Figure 4.11e, with peaks coinciding with discharge troughs. Calcium, magnesium, manganese, and aluminum also show this inverse relationship to discharge (see Figure 4.11f of aluminum for example).

Figure 4.11a: Plot of Discharge vs. Time (Arnot 003)

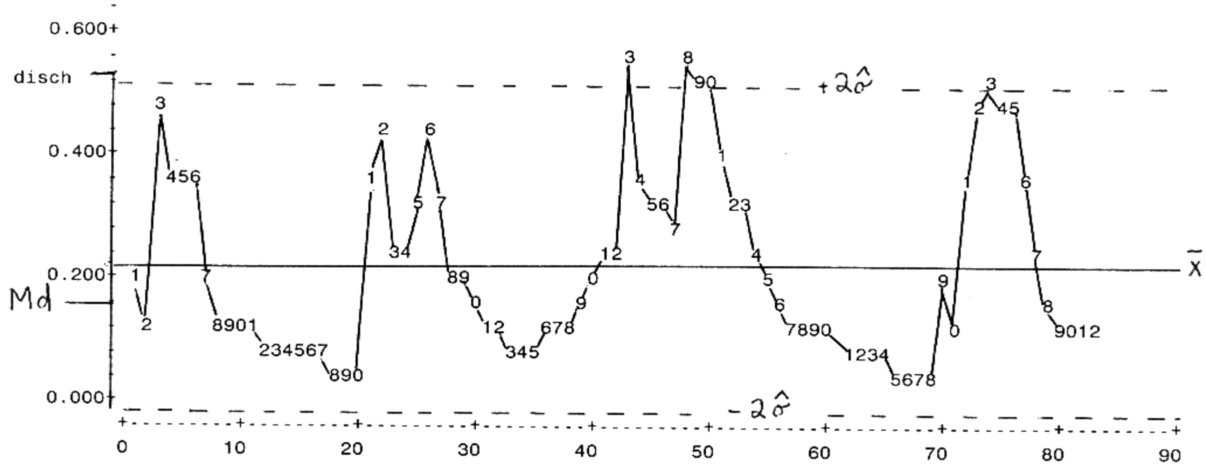


Figure 4.11b: Plot of pH vs. Time (Arnot 003)

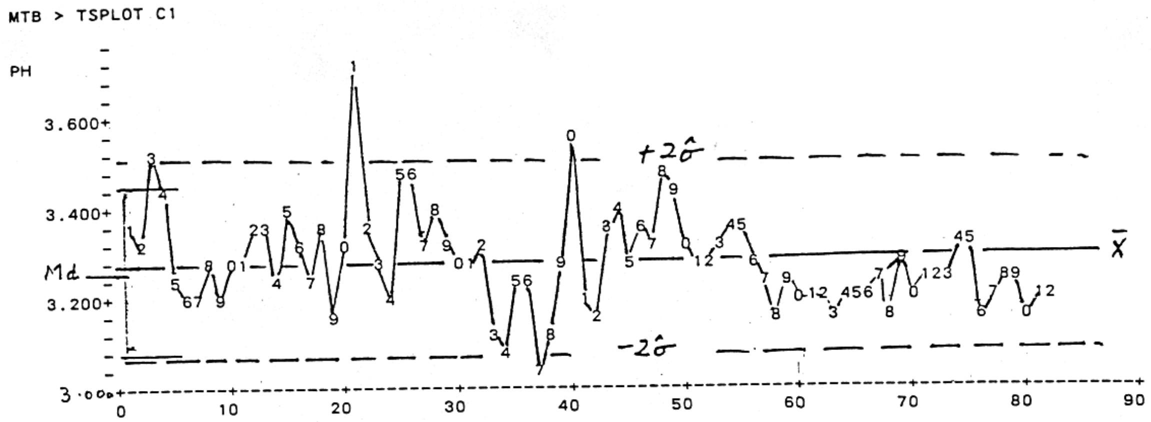


Figure 4.11c: Plot of Acidity vs. Time (Arnot 003)

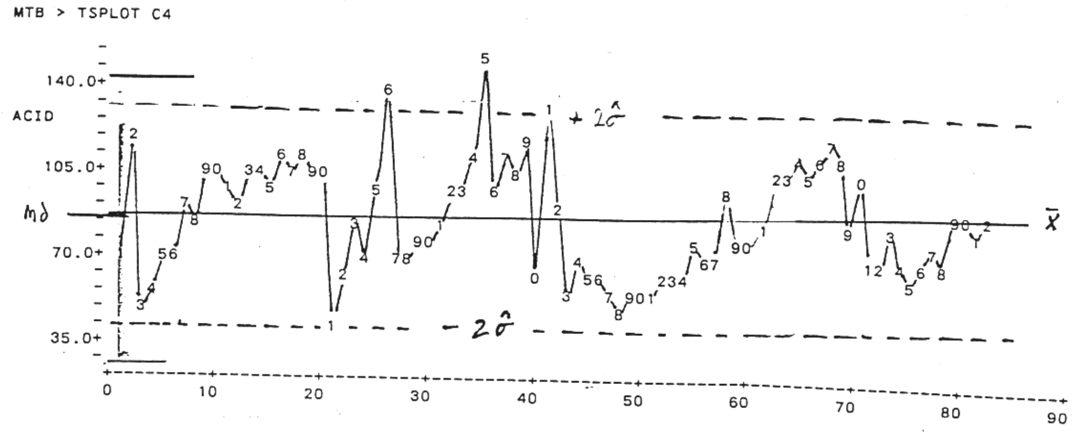


Figure 4.11d: Plot of Total Iron vs. Time (Arnot 003)

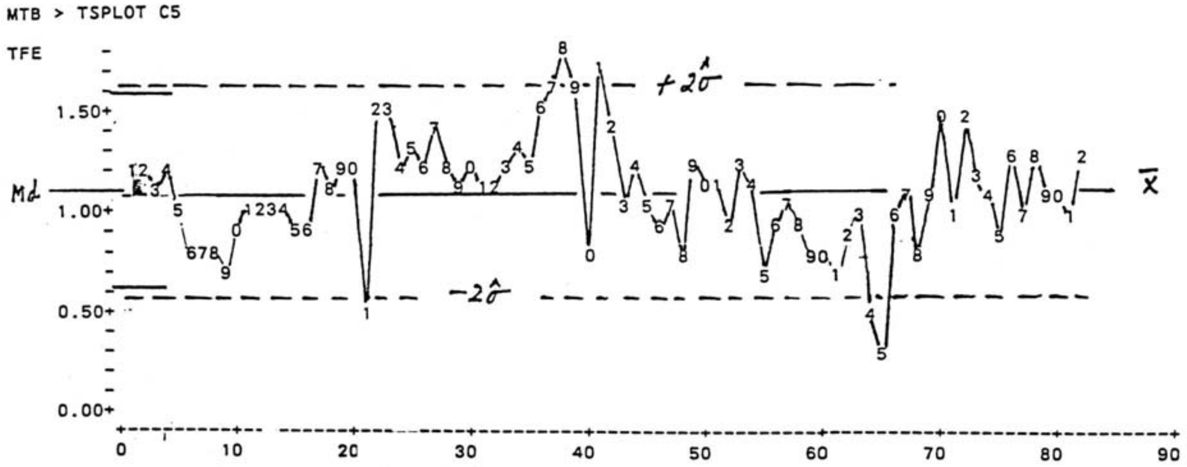


Figure 4.11e: Plot of Sulfate vs. Time (Arnot 003)

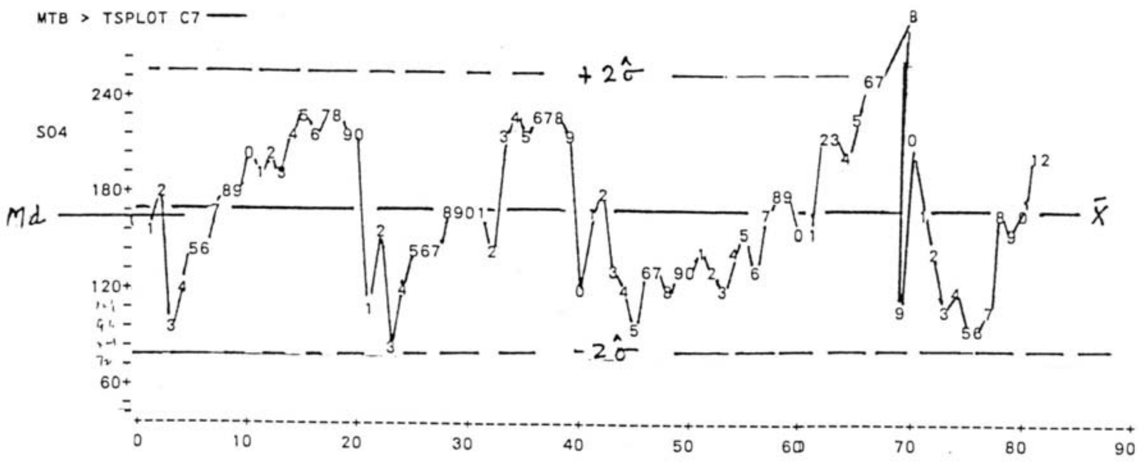
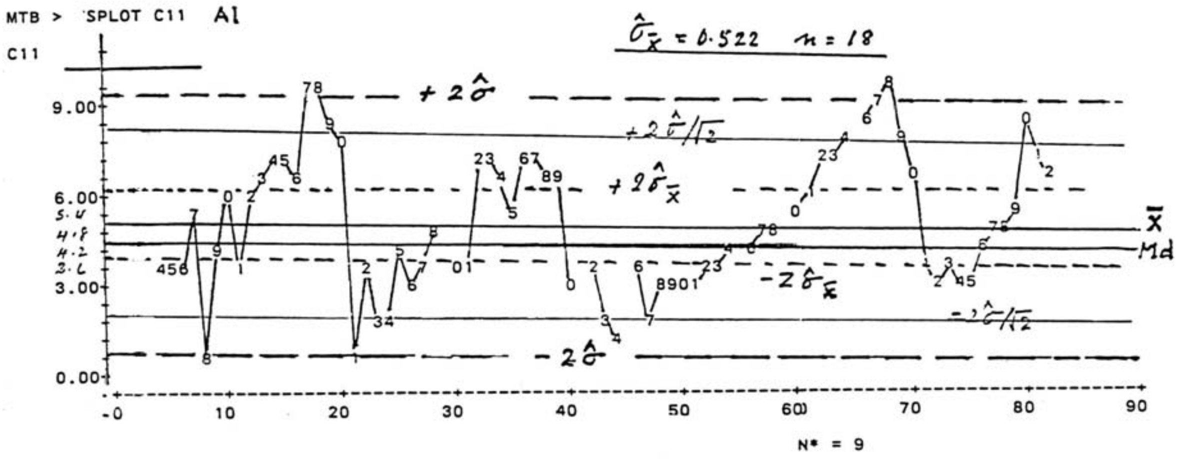


Figure 4.11f: Plot of Aluminum vs. Time (Arnot 003)



The time series plots shown in Figures 4.11a to 4.11f can be used as quality control graphs in the following manner. Confidence limits around the mean are simple to prepare from the descriptive statistics in Tables 4.1a to 4.1c and these can be inserted in, for example, Figure 4.11c. Two kinds of confidence limits are included for comparison. The first is based upon the mean and standard deviation of the normal frequency distribution. The second is based upon the median and other order statistics and is for use in cases where the frequency distribution is not normal (e.g. skewed) or in other non-parametric applications. These two kinds of quality control approaches are discussed in more detail in Chapter 5. The most typical quality control limit is the conventional range of the mean (plus or minus two standard deviations) which, in a normal distribution includes some 95 percent of the observations (i.e., one expects in a moderately long (say > 30) series about 2 – 3 observations outside these limits on either side of the mean). If we wish to relax the requirement of a normal distribution we may use the range encompassed by order statistics, for example Md ± 1.58 (H-spr.), which is approximately equivalent to the conventional measure (Velleman and Hoaglin, 1981, p. 81). The multiplier (2) in the conventional example may be replaced with 3 for a more stringent test in which only 3 in 1000 are expected to fall outside the (3 σ) limits, strictly in a normal distribution. The limits for each of the eleven variables from the Arnot 003 data are displayed in Table 4.3, including the range around the means and around the medians. The range around the mean exceeds that around the median in pH, temperature, ferrous iron, and total iron, whereas the range around the median exceeds that around the mean in the seven other variables. These seven variables show associated variation either directly or inversely so this consistency is to be expected. The reason for the reversal in relationship for the other four may arise from inconsistent occurrence of outliers in the data for these variables. pH is usually symmetrical and probably closely normal; temperature, ferric iron and total iron have very marked peculiarities.

Table 4.3: Comparison of Confidence Belts Around Mean and Median (Arnot 003 Data)

	Mean	Std. Dev.	Median	H-spr.	Lower	Upper	Lower	Upper	Range	
Variable	$\overline{X}$	$\hat{\sigma}$	Md	Q3 - Q1	Around Mean		Around Median		Mean	Median
pH	3.2782	0.1095	3.265	0.1225	3.059	3.497	3.071	3.459	0.438	0.388
Temperature	8.551	0.916	8.6	0.9	6.719	10.383	7.178	10.022	3.664	2.844
Flow	0.2157	0.1509	0.161	0.2272	-0.086	0.518	-0.198	0.52	0.604	0.718
Acidity	86.37	22.55	84.5	36.25	41.27	131.47	27.225	141.775	90.2	114.55
Total Iron	1.0963	0.2843	1.1	0.3	0.528	1.665	0.626	1.574	1.137	0.948
Ferrous Iron	0.361	0.234	0.3	0.2	-0.107	0.829	-0.016	0.616	0.936	0.632
SO_4	168.99	43.79	165	77.25	81.41	256.57	42.945	287.055	175.16	244.11
Ca	59.75	11.69	61	20	36.37	83.13	29.4	92.6	46.76	63.2
Mg	73.6	23	70	34	27.6	119.6	16.28	123.72	92	107.44
Mn	3.203	1.338	2.76	2.31	0.527	5.879	-0.89	6.41	5.352	7.3
Al	5.079	2.213	4.68	3.56	0.653	9.505	-0.945	10.305	8.852	11.25

The mean, median, and their associated ranges are included in Figures 4.11a to 4.11f. The means and medians are reasonably close with the median usually being less than the mean. This suggests that the outliers are on the large side (i.e., positive skewness) and are pulling the mean up more than the median. The seven variables which show associated variation should probably all be log transformed. The pH is already in log units, but temperature and the iron variables are not, on the whole, consistent enough to make any general recommendation. Total iron or any combination of these should be carefully checked because their variation is open to a variety of problematic explanations, and until one can be sure that these measures are meaningful, they should be treated with circumspection.

From the point of view of setting up triggers, either of the ranges around the mean or median would suffice. If the confidence belts were constructed around the mean, then for the Arnot 003 data, the following observations fall on, near or totally outside them, as shown in Table 4.4. Apparently the 2 sigma limits are more sensitive to these deviations and the H-spread usually shows less observations outside the limits; since 2 sigma = about 95% confidence limits, then 2.5 are expected to exceed the upper limit. Three, therefore, is an expected number and needs no reaction. The iron observations are again somewhat inconsistent.

Table 4.4: Observations Falling Beyond Confidence Limits of 2 Standard Deviations Around the Mean Beyond the (1.58*) H-Spread (Arnot 003 Data)

	Number of Observations	
Variable	>2 $\hat{\sigma}$	>(1.58) H-Spread
pH	4	8
Temperature	3	7
Discharge	3	3
Acid	3	1

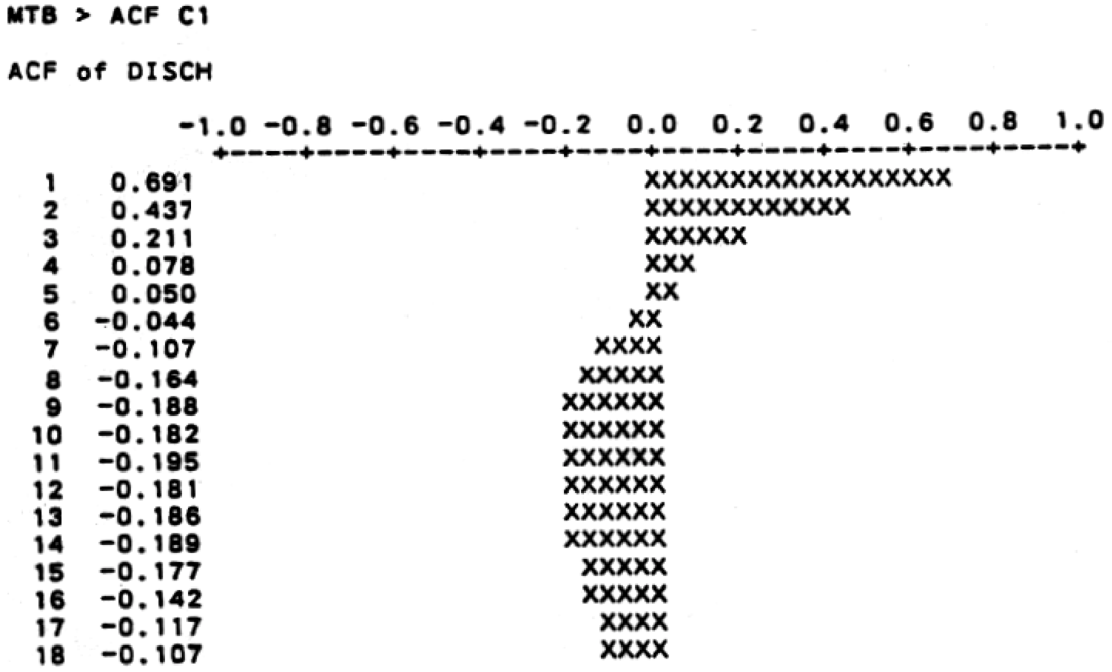
	Number of Observations	
Variable	**>2 $\hat{\sigma}$**	**>(1.58) H-Spread**
Total Iron	6	8
Ferrous Iron	2	8
SO_4	1	0
Ca	2	0
Mg	3	2
Mn	5	0
Al	0	0

The approach to Box-Jenkins Time Series analysis may be simplified to accomplish preliminary exploration of the data. We may, therefore, examine the autocorrelation function (Acf) and the partial autocorrelation function (Pacf) to the data and evaluate their first differences, if necessary. From this analysis it can be decided whether the data appear to represent the Integrated Moving Average (IMA) (0,1,1) model described in Chapter 3, or whether a new model should be fitted.

In general, if the Autocorrelation Factor (Acf) looks more or less J-shaped (e.g., Figure 4.12a for Arnot 001 discharge data), it is close enough to the model already described to need no further analysis. If it is subsequently decided to pursue the analysis to model fitting then the full Box-Jenkins procedures described in Chapter 3 should be undertaken.

For the Arnot 001 data, the Acf for discharge (Figure 4.12a), calcium (Figure 4.12b), and aluminum (Figure 4.12c) all conform to the J-shape and are considered to be adequately modeled by an IMA (0,1,1) model. The total iron (Figure 4.12d) and ferrous iron graphs do not show this form of Acf so would require a more formal analysis. From these Acf's, however, it is suspected that a simple Moving Average (MA) (0,0,1) would be adequate to represent these data. In other words, the data appear to represent a random walk.

Figure 4.12a: Autocorrelation Function of Discharge (Arnot 001)

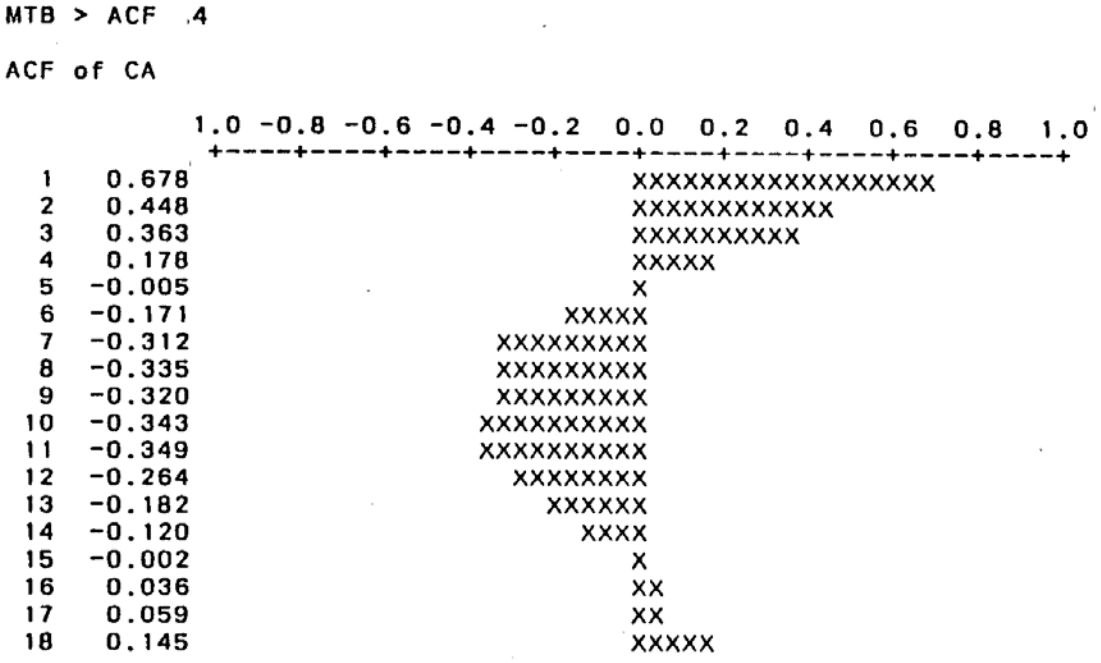


```
MTB > ACF C1

ACF of DISCH

                   -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
                     +----+----+----+----+----+----+----+----+----+----+
   1    0.691                                 XXXXXXXXXXXXXXXXXX
   2    0.437                                 XXXXXXXXXXXX
   3    0.211                                 XXXXXX
   4    0.078                                 XXX
   5    0.050                                 XX
   6   -0.044                                XX
   7   -0.107                              XXXX
   8   -0.164                             XXXXX
   9   -0.188                            XXXXXX
  10   -0.182                            XXXXXX
  11   -0.195                            XXXXXX
  12   -0.181                            XXXXXX
  13   -0.186                            XXXXXX
  14   -0.189                            XXXXXX
  15   -0.177                             XXXXX
  16   -0.142                             XXXXX
  17   -0.117                              XXXX
  18   -0.107                              XXXX
```

Figure 4.12b: Autocorrelation Function of Calcium (Arnot 001)

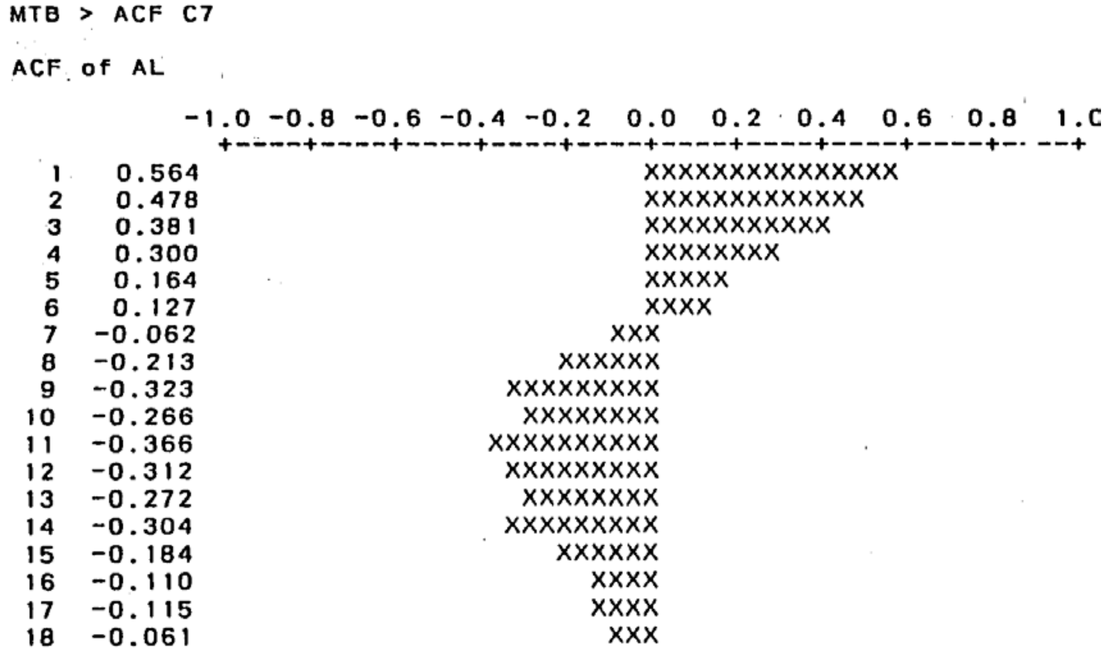


```
MTB > ACF  .4

ACF of CA

                    1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
                     +----+----+----+----+----+----+----+----+----+----+
   1    0.678                                 XXXXXXXXXXXXXXXXXX
   2    0.448                                 XXXXXXXXXXXX
   3    0.363                                 XXXXXXXXXX
   4    0.178                                 XXXXX
   5   -0.005                                 X
   6   -0.171                             XXXXX
   7   -0.312                         XXXXXXXXX
   8   -0.335                         XXXXXXXXX
   9   -0.320                         XXXXXXXXX
  10   -0.343                        XXXXXXXXXX
  11   -0.349                        XXXXXXXXXX
  12   -0.264                          XXXXXXXX
  13   -0.182                            XXXXXX
  14   -0.120                              XXXX
  15   -0.002                                 X
  16    0.036                                 XX
  17    0.059                                 XX
  18    0.145                                 XXXXX
```

Figure 4.12c: Autocorrelation Function of Aluminum (Arnot 001)

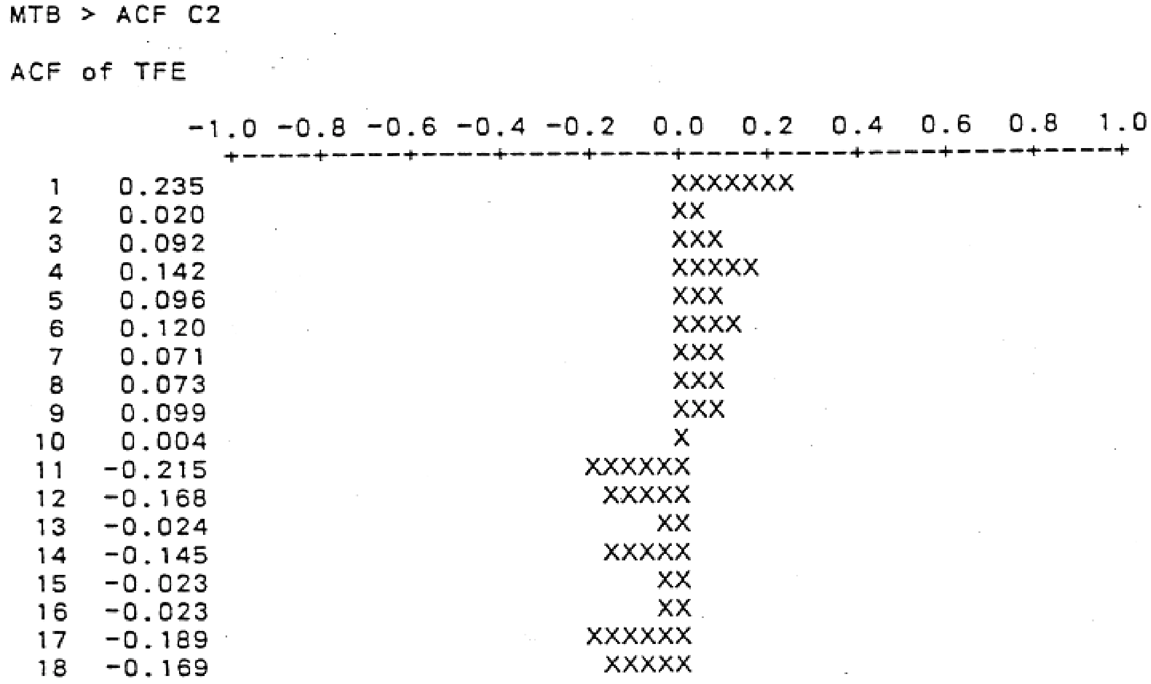


```
MTB > ACF C7

ACF of AL

           -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+- --+
  1   0.564                          XXXXXXXXXXXXXXX
  2   0.478                          XXXXXXXXXXXXX
  3   0.381                          XXXXXXXXXXX
  4   0.300                          XXXXXXXX
  5   0.164                          XXXXX
  6   0.127                          XXXX
  7  -0.062                        XXX
  8  -0.213                     XXXXXX
  9  -0.323                  XXXXXXXXX
 10  -0.266                   XXXXXXXX
 11  -0.366                 XXXXXXXXXX
 12  -0.312                  XXXXXXXXX
 13  -0.272                   XXXXXXXX
 14  -0.304                  XXXXXXXXX
 15  -0.184                     XXXXXX
 16  -0.110                       XXXX
 17  -0.115                       XXXX
 18  -0.061                        XXX
```

Figure 4.12d: Autocorrelation Function of Total Iron (Arnot 001)



```
MTB > ACF C2

ACF of TFE

           -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
  1   0.235                          XXXXXXX
  2   0.020                          XX
  3   0.092                          XXX
  4   0.142                          XXXXX
  5   0.096                          XXX
  6   0.120                          XXXX
  7   0.071                          XXX
  8   0.073                          XXX
  9   0.099                          XXX
 10   0.004                          X
 11  -0.215                     XXXXXX
 12  -0.168                      XXXXX
 13  -0.024                         XX
 14  -0.145                      XXXXX
 15  -0.023                         XX
 16  -0.023                         XX
 17  -0.189                     XXXXXX
 18  -0.169                      XXXXX
```

To check these conclusions, the discharge parameter was run through the full Box-Jenkins autocorrelation function analysis and, as in Chapter 3, a first difference was required to reduce the Acf to that expected from white noise. An autoregressive integrated (ARI) (1,1,0) model was fitted for diagnostic purposes, and while most criteria were satisfactory, the confidence belts around the coefficient of the differenced series included zero. For that reason, this model was rejected and the IMA (0,1,1) appears most appropriate. This analysis of Arnot 001 data was then terminated.

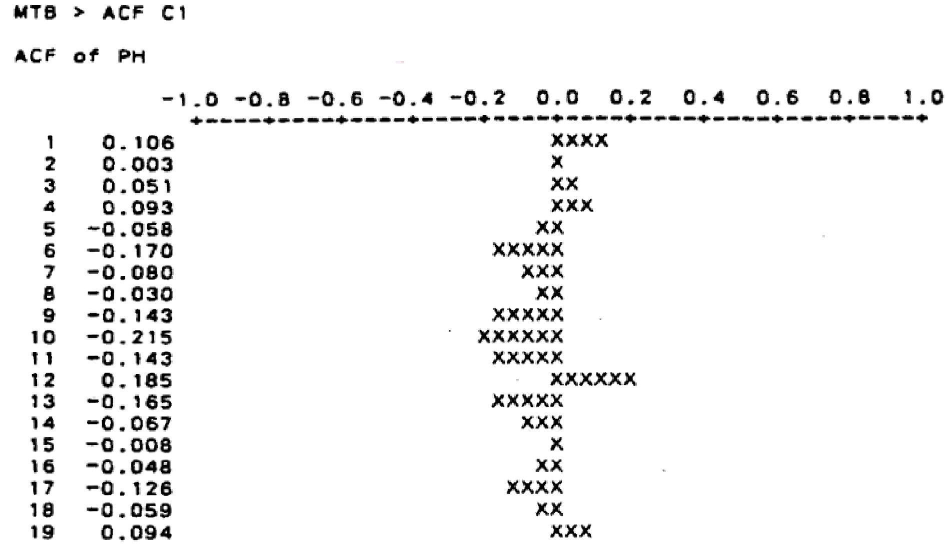
Arnot 003 data yielded similar results and the Acf's of discharge and log discharge were almost identical. Acf's for calcium, magnesium, manganese, and aluminum were similar in form; total iron and ferrous iron are peculiar and probably representative of random variation. A comparison of the standard deviations of the raw data from Table 4.1b and the residuals after fitting the model is illustrated in Table 4.5. There is little improvement from fitting the models, further confirming that the variation in these data are essentially random.

Table 4.5: Comparison of Total Iron and Ferrous Iron

Variable	$\hat{\sigma}$	$\hat{\sigma}_e$
Total Iron	0.284	0.252
Ferrous	0.239	0.231

A few examples of the Acf-Pacf analysis are also included for selected variables from the analysis of the Arnot 004 data. The Acf of pH (Figure 4.13a) is not very informative and the Pacf is identical (Figure 4.13b). Without further analysis these data may be taken to represent a random walk. Log discharge in Figure 4.13c possesses typical features of the IMA (0,1,1) model, a rapid decline in the Acf (J-shape) and a single large spike in the Pacf (Figure 4.13d). These features suggest a first difference followed by a first order moving average model.

Figure 4.13a: Autocorrelation Function of pH (Arnot 004)

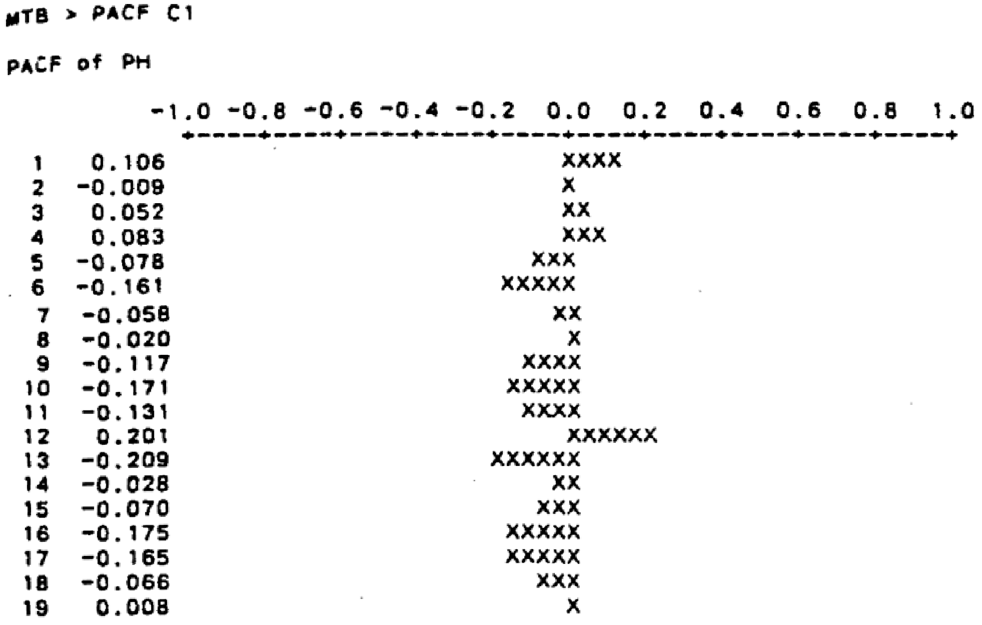


```
MTB > ACF C1

ACF of PH

           -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
             +----+----+----+----+----+----+----+----+----+----+
  1   0.106                           XXXX
  2   0.003                           X
  3   0.051                           XX
  4   0.093                           XXX
  5  -0.058                          XX
  6  -0.170                       XXXXX
  7  -0.080                         XXX
  8  -0.030                          XX
  9  -0.143                       XXXXX
 10  -0.215                      XXXXXX
 11  -0.143                       XXXXX
 12   0.185                           XXXXXX
 13  -0.165                       XXXXX
 14  -0.067                         XXX
 15  -0.008                           X
 16  -0.048                          XX
 17  -0.126                        XXXX
 18  -0.059                          XX
 19   0.094                           XXX
```

Figure 4.13b: Partial Autocorrelation Function of pH (Arnot 004)

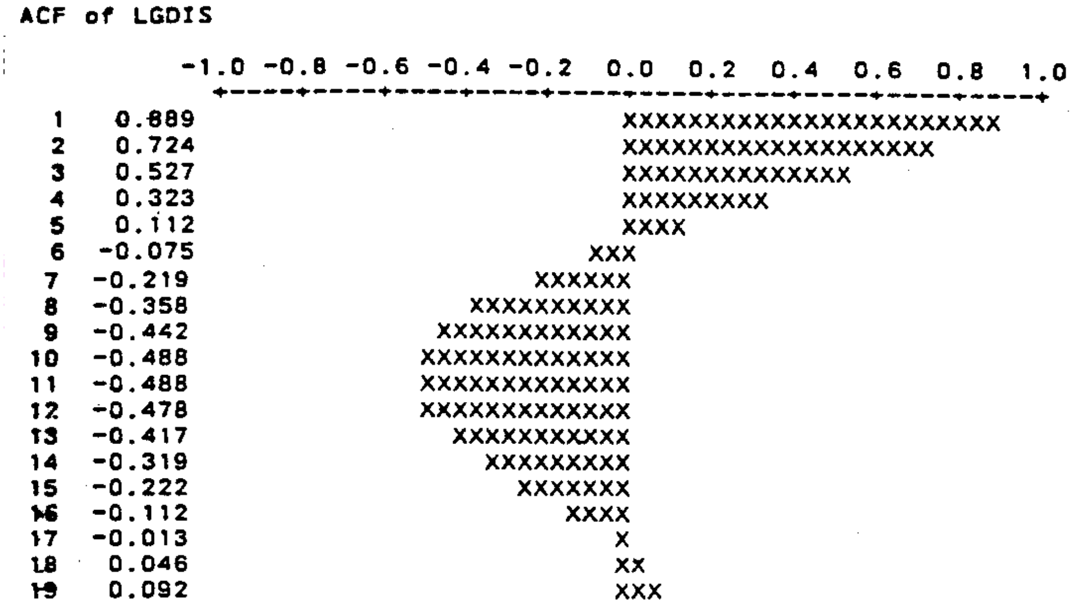


```
MTB > PACF C1

PACF of PH

             -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
              +----+----+----+----+----+----+----+----+----+----+
  1   0.106                            XXXX
  2  -0.009                            X
  3   0.052                            XX
  4   0.083                            XXX
  5  -0.078                          XXX
  6  -0.161                        XXXXX
  7  -0.058                           XX
  8  -0.020                            X
  9  -0.117                         XXXX
 10  -0.171                        XXXXX
 11  -0.131                         XXXX
 12   0.201                            XXXXXX
 13  -0.209                       XXXXXX
 14  -0.028                           XX
 15  -0.070                          XXX
 16  -0.175                        XXXXX
 17  -0.165                        XXXXX
 18  -0.066                          XXX
 19   0.008                            X
```

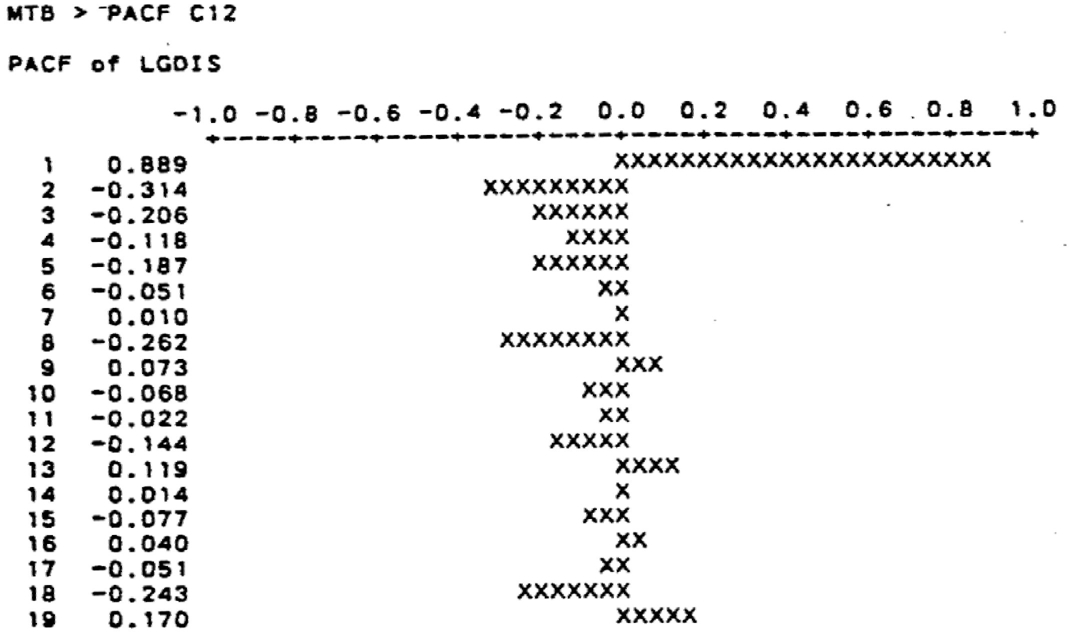
Figure 4.13c: Autocorrelation Function of Log Discharge (Arnot 004)



```
ACF of LGDIS

             -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
              +----+----+----+----+----+----+----+----+----+----+
  1   0.889                            XXXXXXXXXXXXXXXXXXXXXXX
  2   0.724                            XXXXXXXXXXXXXXXXXXX
  3   0.527                            XXXXXXXXXXXXXX
  4   0.323                            XXXXXXXXX
  5   0.112                            XXXX
  6  -0.075                          XXX
  7  -0.219                       XXXXXX
  8  -0.358                   XXXXXXXXXX
  9  -0.442                 XXXXXXXXXXXX
 10  -0.488                XXXXXXXXXXXXX
 11  -0.488                XXXXXXXXXXXXX
 12  -0.478                XXXXXXXXXXXXX
 13  -0.417                  XXXXXXXXXXX
 14  -0.319                    XXXXXXXXX
 15  -0.222                      XXXXXXX
 16  -0.112                         XXXX
 17  -0.013                            X
 18   0.046                            XX
 19   0.092                            XXX
```

Figure 4.13d: Partial Autocorrelation Function of Log Discharge (Arnot 004)



```
MTB > PACF C12

PACF of LGDIS

         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
           +----+----+----+----+----+----+----+----+----+----+
 1   0.889                          XXXXXXXXXXXXXXXXXXXXXXX
 2  -0.314                  XXXXXXXXX
 3  -0.206                     XXXXXX
 4  -0.118                       XXXX
 5  -0.187                     XXXXXX
 6  -0.051                         XX
 7   0.010                          X
 8  -0.262                   XXXXXXXX
 9   0.073                          XXX
10  -0.068                        XXX
11  -0.022                         XX
12  -0.144                      XXXXX
13   0.119                          XXXX
14   0.014                          X
15  -0.077                        XXX
16   0.040                          XX
17  -0.051                         XX
18  -0.243                    XXXXXXX
19   0.170                          XXXXX
```

Figure 4.13e: Autocorrelation Function of Ferric Iron (Arnot 004)

```
         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
           +----+----+----+----+----+----+----+----+----+----+
 1   0.556                          XXXXXXXXXXXXXXX
 2   0.324                          XXXXXXXXX
 3   0.310                          XXXXXXXXX
 4   0.249                          XXXXXXX
 5   0.139                          XXXX
 6  -0.026                         XX
 7  -0.075                        XXX
 8  -0.025                         XX
 9   0.057                          XX
10  -0.001                          X
11  -0.021                         XX
12  -0.069                        XXX
13   0.017                          X
14   0.001                          X
15  -0.120                       XXXX
16  -0.105                       XXXX
17  -0.105                       XXXX
18  -0.120                       XXXX
19  -0.100                        XXX
```

Figure 4.13f: Partial Autocorrelation Function of Ferric Iron (Arnot 004)

```
         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
           +----+----+----+----+----+----+----+----+----+----+
 1   0.556                          XXXXXXXXXXXXXXX
 2   0.023                          XX
 3   0.176                          XXXXX
 4   0.012                          X
 5  -0.051                         XX
 6  -0.181                     XXXXXX
 7  -0.040                         XX
 8   0.057                          XX
 9   0.154                          XXXXX
10  -0.048                         XX
11   0.007                          X
12  -0.163                      XXXXX
13   0.104                          XXXX
14  -0.069                        XXX
15  -0.073                        XXX
16   0.021                          XX
17  -0.055                         XX
18  -0.066                        XXX
19   0.046                          XX
```

Ferric iron shows similar patterns to log discharge, suggesting an IMA (0, 1, 1) model. This is similar to some of the measures of iron content in Arnot 001 and Arnot 003.

Summary

One of the most interesting features in the time series analyses of the Arnot site is the absence or lack of obvious seasonal patterns. Based upon this data set, it appears that this arises for the following reasons:

- The peak flow occurs during Spring snow-melt and runoff. This varies over several months, from February to April, so that successive maxima may not occur at the same time each year.

- Another peak flow may occur in early summer as the result of intense short duration storms. Again this is not strictly confined to exactly the same period from year to year.

- If the missing values occur during these events, and they often appear to be so related, then the extreme values do not occur in a uniform cycle; this confuses any seasonal pattern which may be present.